Function in Engineering: Benchmarking Representations and Models

How to cite:

For guidance on citations see FAQs.

© [not recorded]

Version: Accepted Manuscript

Link(s) to article on publisher’s website:
http://dx.doi.org/doi:10.1017/S0890060417000476

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data policy on reuse of materials please consult the policies page.
Function in Engineering: Benchmarking Representations and Models

Joshua D. Summers (corresponding author)
Professor
203 Engineering Innovation Building - Fluor Daniel
Department of Mechanical Engineering
Clemson University
Clemson, SC 29634-0921
jsummer@clemson.edu

Joshua D. Summers, Professor in Mechanical Engineering at Clemson University, co-directs the CEDAR Group (Clemson Engineering Design Applications and Research). Dr. Summers earned his Ph.D. in Mechanical Engineering from ASU (design automation) and his MS (submarine design) and BS (fluidized bed design) from University of Missouri. He has worked at the Naval Research Laboratory (VR Lab and NCARAI). Dr. Summers’ research has been funded by government, large industry, and small-medium sized enterprises. His areas of interest include collaborative design, knowledge management, and design enabler development with the overall objective of improving design through collaboration and computation.

Claudia Eckert
Professor
Department of Design and Innovation
The Open University
Walton Hall, Milton Keynes
MK7 6AA
UK
Claudia.eckert@open.ac.uk

Claudia Eckert is a Professor of design at Open University. She has a longstanding interest in studying and supporting industrial practice in different design domains and has published numerous papers on it. In particular, she has been working on process modeling, engineering change, and functional modeling of complex engineering products.

Ashok Goel
Professor
School of Interactive Computing
Georgia Institute of Technology
Technology Square Research Building
85 Fifth Street NW
Atlanta, GA 30332
goel@cc.gatech.edu

Ashok K. Goel is a Professor of Computer Science in the School of Interactive Computing at Georgia Institute of Technology in Atlanta, USA. He is the Director of the School’s Ph.D. Program in Human-Centered Computing and the Design & Intelligence Laboratory. He is also Co-Director of the Institute’s Center for Biologically Inspired Design and a Fellow of the Brooke Byer’s Institute for Sustainable Systems. He served on the Board of Directors of The Biomimicry Institute from 2012 to 2017, including as its President from 2015 to 2017. For some thirty years, Ashok has been conducting research into artificial intelligence, cognitive science, and human-centered computing, with a focus on computational design, modeling, and creativity. His 2012 TEDx@Peachtree talk summarizes some of his research on computational design and creativity. An interactive tool for supporting some aspects of biologically inspired design has been developed by his laboratory. Ashok is Editor of AAAI’s AI Magazine and an Associate Editor Design Research Society’s Design Science Journal as well as AIEDAM.
1 ABSTRACT

This paper presents the requirements and needs for establishing a benchmarking protocol that considers representation characteristics, supported cognitive criteria, and enabled reasoning activities for the systematic comparison of function modeling representations. Problem types are defined as reverse engineering, familiar products, novel products, and single-component systems. As different modeling approaches share elements, a comparison of modeling approaches on multiple levels was also undertaken. It is recommended that researchers and developers of function modeling representations collaborate to define a canonically acceptable set of benchmark tests and evaluations so that clear benefits and weaknesses for the disparate collection of approaches can be compared. This paper is written as a call to action for the research community to begin establishing a benchmarking standard protocol for function modeling comparison purposes. This protocol should be refined with input from developers of the competing approaches in an academically open environment. At the same time the benchmarking criteria identified should also serve as a guide for validating a modeling approach or analyzing its failure.

2 FUNCTION REPRESENTATIONS IN ENGINEERING

Reasoning about function of products is critical in product development, which let to many approaches to functional modelling being advocated to support systematic products development. The selection of the approach deemed most appropriate for a particular reasoning need remains problematic, however, given the lack of clear guidance for informing this decision (K. Otto & Wood, 2001; Pahl et al., 2013; Ullman, 2010; Ulrich & Eppinger, 2008). Therefore, this paper provides a justification and a proposed research direction for establishing a common benchmarking scheme for function representations that are developed and deployed throughout academia and practice with the ultimate goal of providing industry with practically usable functional modeling tools and concepts and a clear rationale for selecting a particular one. Despite decades of research into functional descriptions, research suggests industry has yet to incorporate functional modeling in practice in a systematic way, while still proclaiming a need to express product information beyond form (Arlitt, Stone, & Tumer, 2016; Eckert, 2013; Tomiyama, Van Beek, Cabrera, Komoto, & D’Amelio, 2013). It should be noted that there are other representations that are similar, but are not directly associated with function modeling, that are used in industry such as p-diagrams (Campean, Henshall, & Rutter, 2013; Telenko & Seepersad, 2010), block diagramming (Braha & Maimon, 1998; Sturges, O’Shaughnessy, & Kilani, 1996), or IDEF0 (Buede & Miller, 2016; R L Nagel, Perry, Stone, & McAdams, 2009).
Another possible reason for this resistance is the lack of a canonical definition of function with each approach grounded in different conceptualizations, or the possibility of multiple distinct concepts inherent within a shared terminology.

Researchers and practitioners have proposed many different views of function in engineering design (Crilly, 2010; Deng, 2002; Eckert, 2013; Ashok K. Goel, 2013; Srinivasan, Chakrabarti, & Lindemann, 2012; Vermaas, 2013), with three recent approaches included in this special issue on function being the Dimensional Analysis Conceptual Modelling framework (Hossein, Coatanéa, & Henri, 2017), Critical Chain Models (Agyemang, Turner, & Linsey, 2017), and System State Flow Diagram (Yildirim, Campean, & Williams, 2017). These concepts in turn have been used to inform the creation of many approaches for modeling information about a product’s functionality. For example, several design textbooks emphasize the use of function-flow networks to capture the sequence and dependencies for the desired functionality of a product (K. Otto & Wood, 2001; Pahl et al., 2013; Ullman, 2010; Ulrich & Eppinger, 2008).

Rather than develop a single, unified definition of function, we assert that each approach has its own strengths and weaknesses. Although each approach is useful and particularly well suited for different reasoning applications and domains, their transference across different domains remains a difficult proposition. Therefore, we propose a different approach to function research through the creation of a set of comparative benchmarks that can be explored with the different modeling approaches. The community may in turn use these proposed benchmarks to discern which approaches are more useful for different needs, and perhaps to discover which elements of the representations and vocabularies are most conducive for different elements of functional thinking.

The information captured within function models can be used to facilitate many different engineering activities across entire product lifecycle, such as synthesis, analysis, exploration, visualization, explanation, and fault detection (J. Gero & Kannengiesser, 2002; A. K. Goel & Bhatta, 2004; Kurtoglu & Tumer, 2008). While modelling approaches might be defined clearly, they do not always come with clear guidance how to represent specific models. Unfortunately, these modelling approaches and representations are neither perceived as easy-to-use nor to learn with potential users remaining poorly informed with respect to what these representations can provide. Therefore, both of these assumptions represent an educational challenge for the community with the notion of function even considered as unclear (Eckert, Alink, Ruckpaul, & Albers, 2011). Therefore, these assumptions must be addressed with each
representation and modeling approach proposed, but with common frames of reference, perhaps supported through a standard benchmarking protocol that would define common problems and common issues against which methods can be challenged.

3 THE NEED FOR BENCHMARKING

We next turn to other canonical benchmark systems that have proven useful in the cross-comparison of algorithms and methods. Despite the subject of repeated analysis in the research community, benchmarking has rarely been applied to the methods and tools developed in the engineering design community. One of the few examples of an effort to directly compare different function representations focused on additive manufacturing processes (Joshua D. Summers & Rosen, 2013). For example software engineering may entail the use of benchmarking to compare algorithms (Dolan & Moré, 2002), specifically as regards to the Traveling Salesman Problem which involves a series of benchmark problems used for such comparisons (Peterson, 1990). Likewise several accepted optimization benchmark problems have been used to evaluate performance of new algorithms (Brest, Greiner, Boskovic, Mernik, & Zumer, 2006). Similarly, standard benchmark tests have been used in the automotive industry to consider different control strategies (Rajamani, 2012).

Each of these different benchmark sets have been constructed to test new algorithms, either optimization or controls. With developing a benchmark set of problems for comparing function representations, the algorithm, or reasoning dimension must be considered as should the representation and the modeling of the functions; thus defining a critical distinction between the traditional approaches of benchmarking and the approach proposed here.

No protocol for benchmarking functional modeling currently exists, nor has there been any systematic comparison of the expressive power of various models been, despite the individual comparisons available in the literature of such relationships. Indeed a huge variability between individuals and in particular between different modeling approaches exists, as indicated in the experiment on functional descriptions, where different engineers were provided with a product and asked to generate a functional description (Eckert, Ruckpaul, Alink, & Albers, 2012). Consequently, a systematic analysis is needed to determine how different modeling approaches compare in representational expressive power, reasoning inferencing capacity, and modeling ease of use.
4 IS FUNCTION RESEARCH SUFFICIENTLY MATURE?

Before defining a series of benchmark test cases for use in the cross-evaluation of competing function representations, we must first determine if the research field is sufficiently established to warrant such an effort. This is critical for determining both the sufficient need in a plethora of competing approaches and a sufficient population size of researchers willing to use these benchmarks as comparative tools. To this end, we first consider the field’s evolution over the past five decades (Table 1). The evolution has been characterized by incremental shifts and the creation of function vocabularies (from normative function descriptions (Pahl et al., 2013) to controlled vocabularies (J. Hirtz, Stone, McAdams, Szykman, & Wood, 2002) to physics-defined vocabularies (Chiradeep Sen, Summers, & Mocko, 2013c)). It has also involved the creation of conceptually divergent approaches for modeling interface centered (Wang, Qing-liang, & Albers, 2009), component centered (Fenves, Foufou, Bock, & Sriram, 2008), and user centered functionalities (Gaffney, Maier, & Fadel, 2007)). Functional modelling currently receives a renewed interest as the greater integration of mechanical systems with electric systems and software required trade-off across systems on a functional. For example the IFM (Gericke & Eisenbart, 2017) in this special issue bridges this gap by combining multiple views on functional modelling.

Table 1: Recent Decades of Engineering Function Research

<table>
<thead>
<tr>
<th>Decade</th>
<th>Example References</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960’s</td>
<td>(Eastman, 1969; Pahl et al., 2013)</td>
</tr>
<tr>
<td>1970’s</td>
<td>(Collins, Hagan, &amp; Bratt, 1976; Freeman &amp; Newell, 1971; Rodenacker, 1971)</td>
</tr>
<tr>
<td>2000’s</td>
<td>(Albers, Thau, &amp; Alink, 2008; Chandrasekaran, 2005; Erden et al., 2008; J. Gero &amp; Kannengiesser, 2002; J. Hirtz et al., 2002)</td>
</tr>
<tr>
<td>2010’s</td>
<td>(Linz, 2011; Schultz et al., 2010; C Sen, Summers, &amp; Mocko, 2011; Srinivasan et al., 2012; Yang, Patil, &amp; Dutta, 2010)</td>
</tr>
</tbody>
</table>

There are many representations available, such as the Structure-Behavior-Function (Bhatta & Goel, 1997), the Function-Behavior-Structure (Qian & Gero, 1996), the Functional Basis (J. Hirtz et al., 2002), the Function-Behavior-State (Umeda et al., 1996), Affordance Based Design (Maier, Srinivasan, Vikram, Summers, & Bernhard, 2007), the Contact and Channel Model (Albers et al., 2008), and the general function lists used in such design tools as morphological charts (Smith, Richardson, Summers, & Mocko, 2012), the House of Quality.
Developed by many researchers worldwide, each of these function representations and modeling approaches are characterized with a different intent, history, and context behind the representation (Erden et al., 2008). It is because of the many different roles and uses within engineering design that these models have evolved disparately. Unfortunately, many of these approaches are limited by the “inventors” problem within design research – in which researchers will push the creation a solution to a problem and a design need without actually designing the tool or method based on the intrinsic properties. Thus, many of the representations, while serving different specific purposes such as machinery and manufacturing systems with an emphasis on flows (Pahl et al., 2013), might support other activities addressed by competing representations. This discrepancy suggests the need for developing a systematic comparison system.

Most of the modelling approaches were developed in response to a problem that would have been difficult to approach using methods with which the authors were familiar, and that work particularly well on certain problems. For example, although the functional models based on flow of energy, matter, and information work extremely well for production machinery (Pahl et al., 2013), in which many examples are used inform the model, they are ineffective in elucidating the functions a single complex component (e.g. the turbine blade of a jet engine. Further, although most of the functional modelling approaches in the literature use examples to illustrate their points, they neither discuss the scope of their approach nor reflect about the applicability to other classes of problems.

5 LEVELS OF COMPARISON

A thorough analysis this problem must be predicated on a short discussion of the different levels at which function modeling, representation, and reasoning can be compared. Such comparisons may range from the fundamental core idea through to the use of models (Figure 1). The most theoretical level of comparisons involves studying differences of “purpose”, “transformation”, or “intent” (Rosenman & Gero, 1998), which have been reduced to formal frameworks such as the Function-Behavior-Structure (J. S. Gero & Kannengiesser, 2004). Such frameworks have been used to create defined representations for instantiating models, that are then couple with reasoning activities, such as “model building” and “model using” for drawing inferences (Cebrian-Tarrason, Lopez-Montero, & Vidal, 2008). Tools such as the FunctionCAD (R L Nagel et al., 2009), 2nd-CAD (Vargas-Hernandez & Shah, 2004), function-behavior-state modeler (Umeda et al., 1996), design repository (Bohm, Stone, & Szykman, 2005), ConMod (Chiradeep Sen, Summers, & Mocko, 2013b) have been used to
support such model building activities. The models created from these tools have then been used to for different inferences, such as failure modes (Stone, Tumer, & Van Wie, 2005), predicting assembly time (Mohinder, Gill, & Summers, 2017), and predicting assembly time (Gill & Summers, 2016; Mathieson et al., 2011). Thus, we must consider the primary factors of representation and reasoning, and their interaction with the modeler and user.

Figure 1: Level of Comparison and Consideration during Function Benchmarking

Different functional modelling approaches have different core ideas at the core of the approach. For example, some analyze function is terms of the flow (Pahl et al., 2013), whereas other approaches involve elucidating the relationship of function, structure, and behavior (J. S. Gero & Kannengiesser, 2004). Underpinning the core idea of any of these approaches are various notions of functions (Vermaas, 2013).

The different modelling approaches usually are supplied with a modelling formalism, which are the parameters in which a functional model is described. For example, functions both verb-noun pairs (Julie Hirtz, Stone, & McAdams, 2002; R L Nagel et al., 2009; Robert L Nagel, Stone, Hutcheson, McAdams, & Donndelinger, 2008) and either sentences or single words have been used to describe a given function (Deng, 2002). Conversely, contact and channel have been used to prescribe the elements contained within the description of a function. One such formalism involves the description of a with at least two working surface pairs where the function is enacted (Albers et al., 2008)

The expressive power of these functional models is greatly affected by such formalisms that are used in their construction, a variation of great importance when building functional models of complex products, where the effort increases significantly with the complexity of a model. Some modelling approaches are characterized by a hierarchical decomposition of functions while others are not (Erden et al., 2008). The models also vary in the degree of
abstraction required; for example Contact and Channel and function modeling requires a concrete embodiment of the modelling link functions, whereas functional descriptions are abstract in other models. This variation in turn affects the hierarchical decomposition as the lower level description might depend on the chosen embodiment.

The other dimension of comparison is the reasoning with functions both in building models and in using models once they have been generated, which is of importance when the models are used by someone other than their author. In such a case, the familiarity of the modelling approach and the intuitiveness of the representation then becomes a major issue when inferring information. Several studies have been undertaken to study the authorship and consistency and interpretability of models (B W Caldwell, Ramachandran, & Mocko, 2012; Benjamin W. Caldwell, Thomas, Sen, Mocko, & Summers, 2012; Kurfman, Stock, Stone, Rajan, & Wood, 2003), as well as elucidating the correctness of model construction (Robert L Nagel, Bohm, Linsey, & Riggs, 2015). Alternatively, some approaches have been proposed that automatically reasoning on function models from database collections (Lucero, Viswanathan, Linsey, & Turner, 2014; Patel, Andrews, & Summers, 2016; Sridhar, Fazelpour, Gill, & Summers, 2016). Finally, some approaches entail the support of first principle based physics reasoning (Ashok K. Goel, Rugaber, & Vattam, 2009; Chiradeep Sen, Summers, & Mocko, 2011, 2013a). Thus, it is possible that reasoning might be supported through human use and interpretation or through automated reasoning to infer information.

Some functional modelling approaches, such as functional basis (Julie Hirtz et al., 2002), are supported by dedicated modelling tools (Bryant, McAdams, Stone, Kurtoglu, & Campbell, 2006; R L Nagel et al., 2009; Vargas-Hernandez & Shah, 2004), which allow the user to build model more comfortably and therefore also support building larger models. The usability of the tools affects the potential success of the modelling approach independent of the other elements.

6 COMPARISON ACROSS THE CRITERIA

The framework in Figure 1 proposes different criteria on which modelling approaches can be compared. Here we detail our criteria for comparing these criteria by proposing characteristics along which the comparison can be made.

6.1 Representation Characteristics

When comparing function modeling approaches, the typical approach initially involves comparing the representations at the formalism level, including vocabulary and grammar. A
representation is the formalism through which a model is constructed which means that a model is instantiated through a representation. Multiple models of the same “real world” target may be created through the same representation with each distinct model clearly mapped to a single real-world target. Ambiguity is introduced when one model serves as surrogate for multiple targets (Shah & Mantyla, 1995), which are not clearly specified.

Similar comparison criteria, derived from research in artificial intelligence, include representational adequacy, inferential adequacy, inferential efficiency, and acquisitional efficiency (Winston, 2005). Another approach to compare representations examines the vocabulary, structure, expression, purpose, and abstraction (J D Summers & Shah, 2004). Expanding upon that research, we propose that the representation comparison should include, but not limited to:

- **scope**: the domain for which the function modeling approach is intended (R L Nagel, Vucovich, Stone, & McAdams, 2008);
- **flexibility**: the ability to modify and adapt the representation to address new problems (Regli, Hu, Atwood, & Sun, 2000);
- **indexing**: support access to the right (or useful) knowledge when needed (A. K. Goel & Bhatta, 2004)
- **consistency**: enforce physics and other consistencies (Chiradeep Sen et al., 2011);
- **translationability**: tied to other engineering models (Nebel, 2000);
- **behavior**: ability of the representation to simulation behavior (Qian & Gero, 1996);
- **scalability**: support both simple and complex problem types (Chiang, Pennathur, & Mital, 2001).

### 6.2 Modeling Characteristics

In addition to these representational issues, the interaction of the designer during the model construction is of concern when comparing the functional modeling approaches. For instance, is the modeling computationally supported, restricted to human effort, or is a mixed initiative approach supported (Chiradeep Sen et al., 2013a). Additionally, the support of various construction approaches within the model, such as forward chaining (moving from input to output), backward chaining (moving from output to input), nucleation, environment to system (outside to inside), or system to environment (inside to outside) are also important considerations (Chiradeep Sen & Summers, 2014). A final characteristic relates to the support of decomposition and recomposition across multiple hierarchical levels and abstractions within the respective modeling approach (Pahl et al., 2013).
6.3 Cognitive Criteria Characteristics

The concept of cognitive criteria has been developed in Human Computer Interaction to help software designers to think through the usability of the artifacts they were creating, such as programming languages or user interfaces (Green, 1991). Although many developers and software engineers have experience in developing well-design information artifacts, they have no way of articulating why these approaches are appropriate for meeting user needs. Functional modeling approaches are considered information artifacts, similar to programming languages and cognitive criteria offer a vocabulary for discussing usability issues, that is informed by cognitive science (Blackwell et al., 2001). The framework is deliberately broad to avoid being overwhelmed in the details of an implementation and thereby losing the sought-after conceptual improvements. However, this task-specific approach, which addresses processes and activities rather than merely the final product, means that it can be used to evaluate and not simply compare functional modeling approaches. The cognitive criteria are orthogonal in supporting reasoning trade-offs and for analyzing the space of possible solutions in a coherent manner, and where possible observing the effect of combinations of criteria. Table 2 presents a selection of the cognitive criteria with their questions for programming and a possible interpretation of these questions for functional modeling, which would require refinement prior to a benchmarking exercise (Green & Petre, 1996).

Table 2: Key cognitive dimensions (criteria) based on (Green & Petre, 1996)

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Question for Programming Languages</th>
<th>Question of Functional Modelling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstraction Gradient</td>
<td>What are the minimum and maximum levels of abstraction? Can fragments be encapsulated?</td>
<td>What are the minimum and maximum levels of abstraction? Can a partial model be created?</td>
</tr>
<tr>
<td>Closeness of mapping</td>
<td>What ‘programming games’ must be learned?</td>
<td>What modeling conventions must be learned? How intuitive is the resulting model?</td>
</tr>
<tr>
<td>Error-proneness</td>
<td>Does the design of the notation induce ‘careless mistakes’?</td>
<td>Does the design of the notation induce ‘careless mistakes’?</td>
</tr>
<tr>
<td>Hidden dependencies</td>
<td>Is every dependency overtly indicated in both directions? Is the indication perceptual or only symbolic?</td>
<td>Is every dependency overtly indicated in both directions? Is the indication perceptual or only symbolic?</td>
</tr>
<tr>
<td>Premature commitment</td>
<td>Must programmers make decisions before they have the information they need?</td>
<td>Does the model require decisions prior to availability of all necessary information?</td>
</tr>
<tr>
<td>Secondary notation</td>
<td>Can programmers use layout, color, or other cues to convey extra meaning, above and beyond the ‘official’ semantics of the language?</td>
<td>Can the models be annotated or linked to other product representations?</td>
</tr>
</tbody>
</table>
How much effort is required to perform a single change? What is the difficulty of adapting the model from a model of a similar product?

Is every part of the code simultaneously visible (assuming a large enough display), or is it at least possible to juxtapose any two parts side-by-side at will? If the code is dispersed, is it at least possible to know in what order in which to read?

How easy is it to see all elements of the model? Can two models be compared?

The computing cognitive criteria have a dimension of the **progressive evaluation**, which is the method for obtaining feedback on the modeling through the process and which appears far more meaningful for a programming language, which can be deployed in many different ways. Criteria such as **diffuseness**, which address the number of symbols or graphic entities required to express a meaning, and **hard mental operations**, which questions the need for annotations, also specifically address the notion. Both greatly depend upon a particular implementation version, as few standards concerning functional modeling have yet to emerge.

For each of the dimension a scale of sub categories can be developed such as the abstraction gradient, which is decomposed into abstraction-hating, abstraction-tolerant, and abstraction-hungry (Green & Petre, 1996). For example, while abstraction-hungry programming languages may be considered difficult, abstraction can reduce error-proneness and increase viscosity.

### 6.4 Reasoning Characteristics

Reasoning is the comparison dimension that motivates the need for a common, standard benchmark for evaluating function modeling approaches. It is for different classes of reasoning that each function model is constructed. These reasoning activities entail failure detection (Kurtoglu & Tumer, 2008), reverse engineering and product understanding (J. Hirtz et al., 2002), design decision justification (J. Gero, 1996), design verification and validation (Wiltgen & Goel 2015)(Wiltgen & Goel, 2016), or concept definition and exploration (Pahl et al., 2013). Some types of reasoning that can be evaluated with respect to support include:

- **interpretability**: how consistent and precise is the interpretation of the function models across different individuals, domain, and expertise (Benjamin W. Caldwell et al., 2012);
- **physics maintenance**: can questions about conservation of energy or material, irreversibility, or other physics queries be answered (Chiradeep Sen et al., 2011);
• *analogical mapping*: does the representation support analogical mapping and alignment (Qian & Gero, 1996);

• *pattern learning*: does the representation support the learning of abstractions required for analogical transfer (Bhatta & Goel, 1997);

• *state transformations*: does the representation support answering questions about different states (Deng, 2002);

• *change propagation*: does the representation support discovery about the effects of perturbations in the system (Kurtoglu & Tumer, 2008).

These reasoning criteria might relate to the cognitive criteria. For example, interpretability, analogical mapping, and change propagation might relate to closeness of mapping and viscosity. On the other hand, physics maintenance and state transformation is focused more on the content of the model. Other challenges in reasoning might relate to the ability to contextualize the system within a larger environment or the distribution of system level functions to several distributed elements.

7 **BENCHMARK PROBLEM TYPES**

In order to explore the different characteristics of representation that enable cognitive criteria to support reasoning activities, a set of benchmark problems are needed. Expanding upon the problem type classifications in the literature, we propose four types for study, an example for each type found in the literature, and a list of alternative examples for each. We notably do not include large scale, complex systems such as submarines, aircraft, or space systems as these are not readily available to all researchers for benchmarking activities. A limitation of many benchmarking efforts is that the problems selected should be relatively simple so that researchers are not dissuaded from applying their approaches against similar problems. Complex systems might be of interest in benchmarking, but the effort involved in constructing these models might be too great when compared to the value of understanding the differences. This challenge is noted here but is not resolved.

7.1 **Reverse Engineered Products**

Many function modeling approaches have been demonstrated on existing products after dissection and reverse engineering. A repository of commercial products, that have been reversed engineered to understand them has been developed with the function representation serving as the foundation for the information model (Bohm et al., 2005). An advantage of including this type of problem in the benchmarking formalism is that the products exist and
their performance can be measured and evaluated. A reverse engineered product provides a common platform for comparison. An example product that has been used extensively (Hamraz, Caldwell, & Clarkson, 2012; Huang & Jin, 2009; Maier, Ezhilan, Fadel, Summers, & Mocko, 2007) to explore function modeling is the hairdryer (Figure 2 and Figure 3). Other possible products that can be considered are pneumatic impact drivers (a greater number of mechanical components), battery power tools (readily available in multiple variations), vacuum cleaners (for comparison across multiple customer cultural differences), bike lights (simple and inexpensive systems) or the glue gun (simple product with material flow) proposed for this special issue (Mocko et al., 2007; Joshua David Summers, Eckert, & Goel, 2017). The selection of a common product is most important so that the community can standardize their demonstration cases.

![Figure 2: Example Knowledge Types in the Hairdryer (Hamraz et al., 2012)](image)

**Figure 2: Example Knowledge Types in the Hairdryer (Hamraz et al., 2012)**

![Figure 3: Function structure of a hairdryer product stored in the Design Repository (http://repository.designengineeringlab.org/)](image)

**Figure 3: Function structure of a hairdryer product stored in the Design Repository (http://repository.designengineeringlab.org/)**

### 7.2 Familiar Product

Engineers use reverse-engineering and the dissection of products to map existing systems and components to specific functionality, the first step of which involves hypothesizing the internal functioning of a product (K. N. Otto & Wood, 1998). Further, while reverse engineering can test the ability of a representation to model the detailed functionality of an existing system, modeling a familiar without the product in hand can expose the ability to be fluidly and flexibly model the system, as significant backtracking and hierarchical jumping is likely. An example of the results from an experimental exercise to explore how engineers model...
known products is found in Figure 4 (Eckert et al., 2011). Different engineers are likely to model the system in different ways even given the same underlying representation, so the expressive power of modeling approaches can be assessed. Thus, this benchmark product can be used to explore defining characteristics of the representation, such as consistency and repeatability, without the cost of buying products to reverse engineer. Other products considered for reverse engineering are bicycles, gear boxes, or printing machines.

**Figure 4: Example Function Model for a Hydraulic Pump** (Eckert et al., 2011)

### 7.3 Novel Products

Generative forward system design involves the development of new multi-component systems for characteristics not previously addressed: novelty (not yet attempted), system (multiple components), and intentional (design with a purpose). Examples of “problems” that used to benchmark and compare different function modeling approaches include automated omelet makers, hand-cranked pretzel makers, shoe-string tying mechanisms, clothes folding machines in hotel laundries, and a hand-cranked automated burrito maker. Ideally, benchmark examples can be drawn from literature to support the objectivity of the benchmark. For example, the burrito folding system problem has been used previously in the comparing of function lists and function structures in morphological charts (Richardson III, Summers, & Mocko, 2011). Figure 5 illustrates the function lists and structures for the burrito folder that were used in these ideation experiments. Such a benchmark is useful in exploring the degree to which a function representation can be used in understanding novel problems and generating new solutions.

**Figure 5: Function List and Function Structure for Burrito Folder** (Richardson III et al., 2011)
7.4 Single-Component Products

Single-component multi-functional products, such as passive morphing airfoils (Schultz et al., 2010) and speed screws are also useful in as benchmarks (Albers et al., 2008). For example, the design and analysis of a speed screw demonstrated the effective use of the Contact and Channel Model in the design of a single component. Given that not function models can capture the functionality and behaviors associated with single components, however, this scaling ability in both small and large systems must be explored. While the speed screw benchmark example (Figure 6) shows the downward scalability of reverse engineering; the passive, morphing airfoil design (Figure 7) illustrates the downward scalability of forward engineered products. Larger scale systems, such as aircraft are not considered within the benchmarking protocol due to the challenge of general access for the researchers.

Figure 6: States of WS Pairs for the C&C Model for a Speed Screw (Albers et al., 2008)
Figure 7: Function Structure for a Morphing Airfoil (Schultz et al., 2010)

7.5 Problem Characteristics

While these problem types are focused on the “thing” to be modeled within the function representations, other characteristics may also be considered when comparing the problems. This list is not exhaustive and is intended to recommend to the community other criteria to be considered when developing the benchmark problems for experimental studies. Defining these characteristics will enable the development of similar problems to capture and address comparable criteria.

For instance, measures of the problem size, problem connectivity, and problem difficulty are used to compared problem complexity (Joshua D. Summers & Shah, 2010). Similarly, an information metric was used to compare size complexity (i.e. different problems and prototypes) to explore the impact of representation and level of fidelity in assessing requirement satisfaction (Hannah, Joshi, & Summers, 2011). Finally, problem size as measured with a requirement count and number of words was used to compare design problems in a function modeling experiment (Patel, Kramer, Summers, & Shuffler-porter, 2016; Worinkeng, Joshi, & Summers, 2015).

Other considerations used in defining a design problem have included:

- The domain appropriateness for the participants (Ostergaard, Wetmore III, Divekar, Vitali, & Summers, 2005; Chiradeep Sen & Summers, 2014; Wetmore III, Summers, & Greenstein, 2010);
- realistic and not contrived concepts (Linsey et al., 2010; Thiagarajan, Patel, O’Shields, & Summers, 2017);
• constraints and conditions of the problem (Stacey and Eckert, 2010, Eckert and Stacey, 2014)
• a compelling and intrinsic motivation for the participants (Joshi & Summers, 2014; Linsey et al., 2010);
• representation characteristics of both expression and abstraction (Hannah et al., 2011; McKoy, Vargas-Hernández, Summers, & Shah, 2001; J D Summers & Shah, 2004).

The authors wish to emphasize that the problems developed should be common across multiple criteria for the cross-comparison between benchmarking results, particularly when using sampling logic for the experimental analysis.

8 RESEARCH RECOMMENDATIONS

While this paper is written as a call-to-action for the establishment of standards for function modeling benchmarking, such standards should be developed and evolved with input from both users and developers of the various modeling approaches. Such collaborations should also occur in an academically open environment. Researchers will find clear value in this benchmarking process as it (1) forces the disparate communities to begin to communicate with each other, (2) distributes tutorials on the creation and execution of a variety of models and methods to enhance the education of future engineers, and (3) can be paired with a reasoning/representation selection database to systematically develop informed tools and methods. The benchmarking exercises can also help researchers justify a systematic evolution of their approaches.

We do not see benchmarking as an alternative to validation and verification of functional modelling approaches, but as an addition, that can help in the process of evaluating the approach. While validation and verification can attest to the completeness and correctness of a functional modeling approach, benchmarking helps us compare multiple approaches in terms of information captured, reasoning supported, computational efficiency, and cognitive plausibility.

An internationally diverse benchmarking development group representing the various approaches is a recommended strategy for creating this protocol. In addition to the intellectual motivation for benchmarking, this group can pressure both public and private sector agencies to allocate research funds to support the development of benchmarking schemes.

These benchmarking criteria described above are not intended to serve as the final set, but rather to serve as the point of departure. Additional study is needed to refine these
characteristics and to strategically select case examples for comparison. This proposed development group can support these endeavors directly.

9 REFERENCES


Kurtoglu, T., & Tumer, I. (2008). A Graph-Based Fault Identification and Propagation


FIGURES:

Figure 1: Level of Comparison and Consideration during Function Benchmarking

Figure 2: Example Knowledge Types in the Hairdryer (Hamraz et al., 2012)

Figure 3: Function structure of a hairdryer product stored in the Design Repository (http://repository.designengineeringlab.org/)
Figure 4: Example Function Model for a Hydraulic Pump (Eckert et al., 2011)

Figure 5: Function List and Function Structure for Burrito Folder (Richardson III et al., 2011)

Figure 6: States of WS Pairs for the C&C Model for a Speed Screw (Albers et al., 2008)
Figure 7: Function Structure for a Morphing Airfoil (Schultz et al., 2010)